

Desarrollo de un modelo de scoring de horizonte temporal continuo

(Problema planteado por Management Solutions)

Exposición del problema:

Un modelo de scoring es una herramienta estadística que permite predecir si un cliente incumplirá o no sus obligaciones de pago con una operación (por ejemplo, una hipoteca o un préstamo personal). Se trata de un modelo predictivo cuya salida es la probabilidad de impago de la operación (*probability of default, PD*).

Para construirlo, se parte de una muestra histórica que contiene datos de préstamos históricos que han sido morosos o no morosos. Formalmente,

- $\{x_1, x_2, \dots, x_k\}$ son las k variables observadas para cada préstamo (por ejemplo, edad, ingresos, importe solicitado, etc.).
- δ es el indicador de si un préstamo resultó moroso.

Los modelos habituales requieren que el indicador δ de cada préstamo sea una variable binaria: 0 bien resultó moroso (1) o bien resultó no moroso (0).

El objetivo, por tanto, es encontrar f tal que $P(\delta = 1) = f(x_1, x_2, \dots, x_k)$.

Para ello, es necesario **definir un horizonte temporal** en el que la morosidad ocurrió o no ocurrió. Lo habitual es:

- Considerar el evento $\delta = 1$ si el préstamo fue moroso durante su **primer año** de vida, y por tanto el modelo construido predice el impago a un horizonte temporal de un año (PD a un año).
- Considerar el evento $\delta = 1$ si el préstamo fue moroso en **algún momento de su vida**, y por tanto el modelo construido predice el impago a un horizonte temporal igual a la vida del préstamo (PD a toda la vida).

Ambas opciones son limitadas, porque modelizan la PD a un año o a toda la vida. En la práctica, se desea utilizar la PD a horizontes arbitrarios. Por ejemplo:

- ¿Cuál es la probabilidad de que la hipoteca 0, dada por $\{x_1^0, x_2^0, \dots, x_k^0\}$, sea morosa durante su cuarto año de vida?

El problema propuesto consiste precisamente en desarrollar un modelo de scoring que permita calcular probabilidades de impago (PD) a horizontes temporales arbitrarios.

Esquema del trabajo a realizar:

1.1. Ejercicio preliminar: Modelo de scoring tradicional (4 horas)

Como primera parte del problema, puede resultar interesante construir un modelo de scoring “tradicional”, con un horizonte temporal prefijado, para familiarizarse con los datos y asegurarse de comprender bien el objetivo perseguido.

Para ello, se debatirán los diferentes modelos predictivos posibles para este tipo de problemas y finalmente se utilizará una **regresión logística** (logit) o un modelo **probit** para su resolución. Esta fase concluirá con la validación y los comentarios por parte de MS de los resultados del ejercicio realizado.

1.2. Problema: Desarrollar un modelo de horizonte temporal continuo (10 horas)

Una vez concluido el ejercicio, conocidos los datos y la naturaleza del problema principal, se avanzará hacia su resolución, empleando las herramientas matemáticas e informáticas que el problema requiera.

Una opción a priori posible es la **regresión de Cox**, si bien es un problema abierto y todas las alternativas metodológicas que se propongan son interesantes.

.

Profesor coordinador del problema:

Ignacio Villanueva Díez